Chordal Averaging on Flag Manifolds and Its Applications

Nathan Mankovich
Colorado State University

Tolga Birdal
Imperial College London

Abstract

This paper presents a new, provably-convergent algorithm for computing the flag-mean and flag-median of a set of points on a flag manifold under the chordal metric. The flag manifold is a mathematical space consisting of flags, which are sequences of nested subspaces of a vector space that increase in dimension. The flag manifold is a superset of a wide range of known matrix spaces, including Stiefel and Grassmanians, making it a general object that is useful in a wide variety computer vision problems.

To tackle the challenge of computing first order flag statistics, we first transform the problem into one that involves auxiliary variables constrained to the Stiefel manifold. The Stiefel manifold is a space of orthogonal frames, and leveraging the numerical stability and efficiency of Stiefel-manifold optimization enables us to compute the flag-mean effectively. Through a series of experiments, we show the competence of our method in Grassmann and rotation averaging, as well as principal component analysis.

1. Introduction

Subspace analysis is a key workhorse of machine learning since various forms of data and parameter sets admit a compact representation as a subspace of a high-dimensional vector space. Diffusion imaging data [19] or appearance variations of objects (*e.g.* human faces) under variable lighting can be effectively modeled by low dimensional linear spaces [10], while a video as a whole can be modeled as the subspace that spans the observed frames [34].

A large body of the aforementioned approaches leverage the mathematical framework of Grassmanian manifolds thanks to the ease in dealing with the confounding variability in observations [21, 22, 23, 27]. As such, they rely on statistical analysis tools inherently requiring mean or variance estimations on matrix manifolds [14, 15, 34]. Yet, (i) they have been found to be susceptible to outliers, and (ii) while Grassmanians were suitable for analyzing *tall* data where the ambient dimension is much larger than the number of data points, they become less effective when it

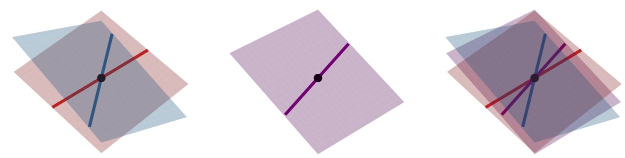

Figure 1: **Chordal averaging** on the **flag manifold** $\mathcal{FL}(1,2;3)$. The average (shown in purple) of the input (red and blue) lines remain in the average of the input planes.

comes to *wide* data where the data dimension is relatively small [29]. In such cases, the more structured *flag manifolds* have been found to be more effective [29].

A flag manifold is a nested series of subspaces geometrically generalizing Grassmanians. Any multilevel, multiresolution, or multiscale phenomena is likely to involve flags, whether implicitly or explicitly. This makes flag manifolds instrumental in dimensionality reduction, clustering, learning deep feature embeddings, visual domain adaptation, deep neural network compression and dataset analysis [35, 29, 50]. Thus, computing statistics on flag manifolds becomes an essential prerequisite powering several downstream applications. In this paper, we propose an approach for computing first order statistics on (*oriented*) flag manifolds (*c.f*. Fig. 1)[1]. In particular, endowing flag manifolds with the non-canonical chordal metric, we first transform the (weighted) *flag-mean* problem into an equivalent minimization on the Stiefel manifold, the space of orthonormal frames, via the method of Lagrange multipliers. We then leverage Riemannian Trust-Region (RTR) optimizers [12, 11] to obtain the solution. Subsequently, we introduce an iteratively reweighted least squares (IRLS) scheme to estimate the more robust *flag-median* as an L_1 flag-mean. Finally, we show how several common problems in computer vision such as motion averaging, can be translated onto averages on flag manifolds using group contraction operators [43]. In particular, our contributions are:

- We introduce a new algorithm for computing flag-prototypes (*e.g.* flag-mean and -median) of a set of points lying on the flag-manifold.
- Analogous to our flag-mean, we introduce an IRLS mini-

[1] While our averages are for general flag-manifolds, we do provide oriented averages for flag manifolds of type $1, 2, 3, \ldots, d-1$ in d-D space.

mization to estimate the flag-median.

- We prove the convergence of the proposed IRLS algorithm for the flag-median.
- We show how rigid motions can be embedded into flags and thus provide a new way to robustly average motions.

Our diverse experiments reveal that flag averages are more robust, usually yield more reliable estimates, and are more general, *i.e*., generalize Grassmannian averages. We will release our implementations upon publication.

2. Related Work

Flag manifolds. Besides being mathematically interesting objects [47, 17, 7], flags and flag manifolds have been explored by a series of works from Nishimori *et al*. addressing subspace independent component analysis (ICA) via Riemannian optimization [39, 41, 38, 42, 38, 40]. Nested sequences of subspaces (e.g. flags) appear in the weights in principal component analysis (PCA) [49] and the result of a wavelet transform [25].

Flag manifolds in computer vision. The utilization of flag manifolds in computer vision is a recent development. Ma *et al*. [29] employ nested subspace methods to compare large datasets. Additionally, they port self-organizing mappings to work on flag manifolds, enabling parameterization of a set of flags of a fixed type. This method was applied to hyper-spectral image data analysis [30]. Ye *et al*. [49] derive closed-form analytic expressions for the set of operators required for Riemannian optimization algorithms on the flag manifold, while Nguyen [37] provides algorithms for logarithmic maps and geodesics on flag manifolds. Marrinan *et al*. [34] investigate the averaging of Grassmanians into flags, demonstrating that flag means behave more like medians and are therefore more robust to the presence of outliers among the subspaces being averaged. Building on this work, they utilize flag averages to improve the detection of chemical plumes in hyperspectral videos [33]. Finally, Mankovich *et al*. [31] also average Grassmannians into flags by providing the median as a flag and an algorithm to compute it.

3. Chordal Centroids on Flag Manifolds

We begin by providing the necessary definitions related to flag manifolds before presenting our chordal flag-mean and -median algorithms.

Definition 1 (Matrix groups)**.** *The* ***orthogonal group*** $O(d)$ *denotes the group of distance-preserving transformations of a Euclidean space of dimension* d*.* $SO(d)$ *is the* ***special orthogonal group*** *containing matrices in* $O(d)$ *determinant* 1*. The* ***Stiefel manifold*** $St(k, d)$*, a.k.a. the set of all orthonormal* k*-frames in* $\mathbb{R}^d$*, can be represented as the quotient group:* $St(k, d) = O(d)/O(d - k)$*. A point on the Stiefel manifold is parameterized by a tall-skinny* $d \times k$ *real matrix with orthonormal columns. The* ***Grassmannian***, $Gr(k, d)$*, represents the collection of points parameterizing the* k*-dimensional subspaces of a fixed* d*-dimensional vector space,* e.g. $\mathbb{R}^d$*. For our purposes,* $Gr(k, d)$ *is a* real matrix manifold*, where each point is identified with an* equivalence class of orthogonal matrices*,* i.e. $Gr(k, d) = O(d)/O(k) \times O(d - k)$*.*
Notation: *We represent* $[\mathbf{X}] \in Gr(k, d)$ *using the truncated orthogonal matrix* $\mathbf{X} \in \mathbb{R}^{d\times k}$*. For this paper* $[\mathbf{X}]$ *is used to denote the subspace spanned by the columns of* $\mathbf{X}$*.*

Definition 2 (Flag)**.** *A* flag *in a finite dimensional vector space* $\mathcal{V}$ *over a field is a sequence of nested subspaces with* increasing *dimension, each containing its predecessor,* i.e. *the filtration:* $\{\emptyset\} = \mathcal{V}_0 \subset \mathcal{V}_1 \subset \cdots \subset \mathcal{V}_k \subset \mathcal{V}$ *with* $0 = d_0 < d_1 < \cdots < d_k < d_{k+1} = d$ *where* $\dim\mathcal{V}_i = d_i$ *and* $\dim\mathcal{V} = d$*. We say this flag is of* type *or* signature $(d_1, \dots, d_k, d)$*. A flag is called* complete *if* $d_i = i$, $\forall i$*. Otherwise the flag is* incomplete *or* partial*.*
Notation*: A flag,* $[\![\mathbf{X}]\!]$ *of type* $(d_1, \dots, d_k, d)$*, is represented by a truncated orthogonal matrix* $\mathbf{X} \in \mathbb{R}^{d\times d_k}$*. Let* $m_j = d_j - d_{j-1}$ *for* $j = 1, 2, \dots, k+1$*, and* $\mathbf{X}_j \in \mathbb{R}^{d\times m_j}$ *for* $j = 1, 2, \dots, k$ *whose columns are the* $d_{j-1} + 1$ *to* d_j *columns of* $\mathbf{X}$*.* $[\![\mathbf{X}]\!]$ *is*

$$[\mathbf{X}_1] \subset [\mathbf{X}_1, \mathbf{X}_2] \subset \cdots \subset [\mathbf{X}_1, \dots, \mathbf{X}_k] = [\mathbf{X}] \subset \mathbb{R}^d.$$

Definition 3 (Flag manifold)**.** *The aggregate of all flags of the same type,* i.e. *a certain collection of ordered sets of vector subspaces, admit the structure of manifolds. We refer to this* flag manifold *as* $\mathcal{FL}(d_1, ..., d_k; d)$ *or equivalently as* $\mathcal{FL}(d + 1)$[2]*. The points of* $\mathcal{FL}(d + 1)$ *parameterize all flags of type* $(d_1, ..., d_k, d)$*. Flag manifolds generalize Grassmannians because* $\mathcal{FL}(k; d) = \mathrm{Gr}(k, d)$*.* $\mathcal{FL}(d + 1)$ *can be thought of as a quotient of groups [30]:*

$$\mathcal{FL}(d+1) = SO(d)/S(O(m_1)\times O(m_2)\times\cdots\times O(m_{k+1})).$$

Definition 4 (Chordal distance on the flag manifold [44])**.** *For* $[\![\mathbf{X}]\!], [\![\mathbf{Y}]\!] \in \mathcal{FL}(d + 1)$*, the* chordal distance *is a map* $d_c : \mathcal{FL}(d + 1) \times \mathcal{FL}(d + 1) \to \mathbb{R}$*:*

$$d_c([\![\mathbf{X}]\!], [\![\mathbf{Y}]\!]) := \sqrt{\sum_{j=1}^{k} m_j - \mathrm{tr}(\mathbf{X}_j^\top \mathbf{Y}_j \mathbf{Y}_j^\top \mathbf{X}_j)}. \quad (1)$$

We now endow flags with *orientation*, which is required in certain applications such as motion averaging.

Definition 5 (Oriented flag manifold [45, 30])**.** *An* oriented flag manifold*,* $\mathcal{FL}^+(d + 1)$*, contains only flags with subspaces with compatible orientations. Algebraically:*

$$\mathcal{FL}^+(d + 1) = SO(d)/(SO(m_1) \times \cdots \times SO(m_{k+1})).$$

[2] Note that we will use $\mathcal{FL}(d_1, ..., d_k; d)$ and $\mathcal{FL}(d+1)$ interchangeably in the rest of the manuscript.

Two oriented vector spaces have the same orientation if the determinant of the unique linear transformation between them is positive [6].

3.1. The Chordal Flag-mean

Armed with notation for flags (Dfn. 2) and ways to measure distance between them (Dfn. 4), we are prepared to state the chordal flag-mean estimation problem formally.

Definition 6 (Weighted chordal flag-mean)**.** *Let $\{[\![\mathbf{X}^{(i)}]\!]\}_{i=1}^{p} \subseteq \mathcal{FL}(d+1)$ be a set of points on a flag manifold with weights $\{\alpha_i\}_{i=1}^{p} \subset \mathbb{R}$ where $\alpha_i \geq 0$. The chordal flag-mean $[\![\boldsymbol{\mu}]\!]$ of these points solves:*

$$\underset{[\![\mathbf{Y}]\!] \in \mathcal{FL}(d+1)}{\arg\min} \sum_{i=1}^{p} \alpha_i d_c([\![\mathbf{X}^{(i)}]\!], [\![\mathbf{Y}]\!])^2. \tag{2}$$

Note: for $\mathcal{FL}(k; n)$, this amounts to the Grassmannian-mean by Draper et al. *[18].*

Proposition 1. *The chordal flag-mean optimization problem in Eq. 2 can be phrased as the Stiefel manifold optimization problem:*

$$\underset{\mathbf{Y} \in St(d_k, d)}{\arg\min} \sum_{j=1}^{k} m_j - \operatorname{tr}\left(\mathbf{I}_j \mathbf{Y}^\top \mathbf{P}_j \mathbf{Y}\right). \tag{3}$$

where the matrices $\mathbf{I}_j$ and $\mathbf{P_j}$ are given below

$$(\mathbf{I}_j)_{i,l} = \begin{cases} 1, & i = l \in \{d_{j-1}+1, d_{j-1}+2, \dots, d_j\} \\ 0, & \textit{otherwise} \end{cases},$$

$$\mathbf{P}_j = \sum_{i=1}^{p} \alpha_j \mathbf{X}_j^{(i)} \mathbf{X}_j^{(i)\top}. \tag{4}$$

Proof sketch. We use truncated orthogonal representations for points on the Stiefel and flag manifolds. By the equivalence of minimization problems we write Eq. 2 as

$$\underset{\mathbf{Y} \in St(d_k, d)}{\arg\min} \sum_{j=1}^{k} m_j - \sum_{j=1}^{k} \sum_{i=1}^{p} \alpha_i \operatorname{tr}\left(\mathbf{Y}_j^\top \mathbf{X}_j^{(i)} \mathbf{X}_j^{(i)\top} \mathbf{Y}_j\right).$$

$\mathbf{I}_j$ allows us to write $\mathbf{Y}_j \mathbf{Y}_j^\top = \mathbf{Y} \mathbf{I}_j \mathbf{Y}^\top$. Using this, properties of trace, and our definition of $\mathbf{P}_j$ we write Eq. 2 as Eq. 3.

□

We provide the full proof in the appendix. We now extend the chordal mean to the case of a certain family of *complete* and *oriented flags*.

Proposition 2. *Let $\{\mathbf{x}^{(i)}\}_{i=1}^{p} \subset \mathbb{R}^d$. Then suppose $\mathbf{x}^{(i)\top} \mathbf{x}^{(j)} > 0$ for all i, j. Then the naive Euclidean mean $\boldsymbol{z} = \frac{1}{n} \sum_{i=1}^{p} \mathbf{x}^{(i)}$ has the same orientation as each $\mathbf{x}^{(i)}$.*

Algorithm 1: Chordal flag-mean.
Input: Set of points on a flag manifold $\{[\![\mathbf{X}^{(i)}]\!]\}_{i=1}^{p}$
Output: Chordal flag-mean $[\![\boldsymbol{\mu}]\!]$
Initialize $[\![\boldsymbol{\mu}]\!]$
Compute projections $\{\mathbf{P}_i\}_{i=1}^{k}$ as in Eq. 4
Define $\{\mathbf{I}_i\}_{i=1}^{k}$ as in Eq. (4)
Optimize Eq. (3) using Stiefel RTR to find $[\![\boldsymbol{\mu}]\!]$

Proof. The proof follows from the simple derivation:

$$\mathbf{x}^{(j)\top} \boldsymbol{z} = \mathbf{x}^{(j)\top} \frac{1}{n} \sum_{i=1}^{p} \mathbf{x}^{(i)} = \frac{1}{n} \sum_{i=1}^{p} \mathbf{x}^{(j)\top} \mathbf{x}^{(i)} > 0.$$

□

Definition 7 ($\mathcal{FL}^+(1, \dots, d-1; d)$ chordal flag-mean)**.** *Let $\{[\![\mathbf{X}^{(i)}]\!]\}_{i=1}^{p} \subset \mathcal{FL}(1, 2, \dots, d-1; d)$ where for each j and any i and k, $\mathbf{X}_j^{(i)\top} \mathbf{X}_j^{(k)} > 0$. Let $[\![\boldsymbol{\mu}]\!]$ be the chordal flag-mean (e.g., Eq. 2) and $\boldsymbol{z}_j$ be the Euclidean mean of $\{\mathbf{X}_j^{(i)}\}_{i=1}^{p} \in \mathbb{R}^d$. Then the oriented chordal flag-mean is defined as $[\![\boldsymbol{\mu}^+]\!] \in \mathcal{FL}^+(1, \dots, d-1; d)^+$:*

$$\boldsymbol{\mu}_j^+ = \begin{cases} \mathbf{Y}_j, & \boldsymbol{z}_j^\top \mathbf{Y}_j \geq 0 \\ -\mathbf{Y}_j, & \textit{otherwise}. \end{cases} \tag{5}$$

Remark 1. *The ordering of the columns of $\boldsymbol{\mu}$ is the same as that of each $\mathbf{X}^{(i)}$ because the chordal distance on the flag manifold respects the ordering of the vectors in the flag representation by only comparing $\boldsymbol{\mu}_j$ to $\mathbf{X}_j^{(i)}$. So, we only need to correct for the sign of the columns of $\boldsymbol{\mu}$. By Prop. 2, we know that the Euclidean mean, $\boldsymbol{z}$, has the same orientation as each of $\mathbf{X}_j^{(i)}$. We use Eq. 5 to force $\boldsymbol{z}_j^\top \boldsymbol{\mu}_j^* \geq 0$. Dfn. 7 gives us a way to choose which chordal flag-mean representatives are best for averaging representations of motions in $\mathcal{FL}^+(1, 2, 3; 4)$ in Sec. 4.*

To compute the proposed mean, we optimize Eq. 3 via RTR methods [2, 12] and re-orient the mean using Dfn. 7.

Remark 2. *The geodesic distance averages on the Grassmannian (e.g. ℓ_2-median and Karcher mean) are known to be unique only for certain subsets of the Grassmannian [3]. The proof of this revolves around finding the region of convexity of the geodesic distance function and its square. Uniqueness for Grassmannian chordal distance averages (e.g. the GR-mean [18] and -median [31]) is largely unstudied. It is known that the chordal distance on the Grassmannian approximates the geodesic distance, but its region of convexity is an open problem to the best of our knowledge. Determining the convexity of our chordal flag-mean and -median would boil down to finding the region of convexity of the chordal distance function and its square on the* *flag manifold. Additionally, one could generalize geodesic distance averages to the flag manifold using Riemannian operators on flags [49], find an algorithm to compute them and their region of convexity. We leave these projects to future work.*

3.2. The Chordal Flag-median

We are now ready to provide our iterative algorithm for robust centroid estimation.

Definition 8 (Weighted chordal flag-median)**.** *Let $\{[\![\mathbf{X}^{(i)}]\!]\}_{i=1}^p \subseteq \mathcal{FL}(d+1)$ be a set of points on a flag manifold with weights $\{\alpha_i\}_{i=1}^p \subset \mathbb{R}$ where $\alpha_i \geq 0$. The chordal flag-median, $[\![\boldsymbol{\eta}]\!]$, of these points solves*

$$\underset{[\![\mathbf{Y}]\!] \in \mathcal{FL}(d+1)}{\arg\min} \sum_{i=1}^p \alpha_i d_c([\![\mathbf{X}^{(i)}]\!], [\![\mathbf{Y}]\!]). \tag{6}$$

Note: for $\mathcal{FL}(k;n)$, this amounts to the Grassmannian-median by Mankovich et al. *[31].*

Proposition 3. *The flag-median optimization problem in Eq. 6 can be phrased with weights $w_i([\![\mathbf{Y}]\!])$ in:*

$$w_i([\![\mathbf{Y}]\!]) = \frac{\alpha_i}{\max\{d_c([\![\mathbf{X}^{(i)}]\!], [\![\mathbf{Y}]\!]), \epsilon\}}, \tag{7}$$

$$\underset{[\![\mathbf{Y}]\!] \in \mathcal{FL}(d+1)}{\arg\min} \sum_{i=1}^p \sum_{j=1}^k m_j - w_i([\![\mathbf{Y}]\!]) \mathrm{tr}\left(\mathbf{Y}_j^\top \mathbf{X}_j^{(i)} {\mathbf{X}_j^{(i)}}^\top \mathbf{Y}_j \right). \tag{8}$$

where $\epsilon = 0$ as long as $d_c([\![\mathbf{X}^{(i)}]\!], [\![\mathbf{Y}]\!]) \neq 0$ for all i.

Proof sketch. We can encode the constraints and our optimization problem into the Lagrangian:

$$\nabla_{\mathbf{Y}_j} \mathcal{L} = -2 \sum_{i=1}^p \frac{\alpha_i \mathbf{X}_j^{(i)} {\mathbf{X}_j^{(i)}}^\top \mathbf{Y}_j}{\sqrt{\sum_{j=1}^k m_j - \mathrm{tr}\left({\mathbf{X}_j^{(i)}}^\top \mathbf{Y}_j \mathbf{Y}_j^\top \mathbf{X}_j^{(i)} \right)}} + 2 \sum_{j=1}^k \lambda_{i,j} \mathbf{Y}_i \mathbf{Y}_i^\top \mathbf{Y}_j,$$

$$\nabla_{\lambda_{i,j}} \mathcal{L} = m_j \delta_{i,j} - \mathrm{tr}\left(\mathbf{Y}_i^\top \mathbf{Y}_j \mathbf{Y}_j^\top \mathbf{Y}_i \right).$$

Then we take the gradient of the Lagrangian with respect to $\mathbf{Y}_j$ and $\lambda_{i,j}$ and set it equal to zero. So, for each j, we have

$$4 m_j \lambda_{j,j} = \sum_{i=1}^p \frac{\alpha_i \mathrm{tr}\left(\mathbf{Y}_j^\top \mathbf{X}_j^{(i)} {\mathbf{X}_j^{(i)}}^\top \mathbf{Y}_j \right)}{d_c([\![\mathbf{X}^{(i)}]\!], [\![\mathbf{Y}]\!])}.$$

Maximizing each $4m_j\lambda_{j,j}$ will minimize the objective function in Eq. 6. We use equivalences of optimization problems to reformulate this maximization as Eq. 8. □

Algorithm 2: Chordal flag-median.

Input: Set of points on a flag manifold $\{[\![\mathbf{X}^{(i)}]\!]\}_{i=1}^p$
Output: Chordal flag-median $[\![\boldsymbol{\eta}]\!]$

Initialize $[\![\boldsymbol{\eta}]\!]$
while *(not converged)* **do**
  Assign $w_i([\![\boldsymbol{\eta}]\!])$ using Eq. (7) (with $\epsilon > 0$)
  $[\![\boldsymbol{\eta}]\!] \leftarrow$ flag-mean$(\{[\![\mathbf{X}^{(i)}]\!]\}, \{w_i([\![\boldsymbol{\eta}]\!])\})$

Proposition 4. *Fixing $[\![\mathbf{Z}]\!] \in \mathcal{FL}(d+1)$, Eq. 8, with $w_i([\![\mathbf{Z}]\!])$, becomes*

$$\underset{[\![\mathbf{Y}]\!] \in \mathcal{FL}(d+1)}{\arg\min} \sum_{i=1}^p \sum_{j=1}^k m_j - w_i([\![\mathbf{Z}]\!]) \mathrm{tr}\left(\mathbf{Y}_j^\top \mathbf{X}_j^{(i)} {\mathbf{X}_j^{(i)}}^\top \mathbf{Y}_j \right)$$

and is equivalent to a chordal flag-mean with weights $w_i([\![\mathbf{Z}]\!])$. Note: $\epsilon = 0$ as long as $d_c([\![\mathbf{X}^{(i)}]\!], [\![\mathbf{Z}]\!]) \neq 0$ for all i.

Proof sketch. This follows from the proof of Prop. 1. □

Prop. 3 simplifies our optimization problem to Eq. 8. Given an estimate for the chordal flag-median, $[\![\mathbf{Z}]\!]$, Prop. 4 shows that solving a weighted chordal flag mean problem will approximate the solution to Eq. 8. Using the propositions, we are now ready to present our iterative algorithm for flag-median estimation in Alg. 2.

The convergence of Weiszfeld-type algorithms are well studied in the literature [4, 8, 51] and our IRLS algorithm for the chordal flag-median can be proven to decrease its respective objective function value over iterations. This is what we establish next in Prop. 5, inspired by the proof methods given in [8].

Proposition 5. *Let $[\![\mathbf{Y}]\!] \in \mathcal{FL}(d+1)$. Suppose $d([\![\mathbf{Y}]\!], [\![\mathbf{X}^{(i)}]\!]) > \epsilon$ for $i = 1, 2, \dots, p$. Also define the maps: $T : \mathcal{FL}(d+1) \to \mathcal{FL}(d+1)$ as an iteration of Alg. 2 and $f : \mathcal{FL}(d+1) \to \mathbb{R}$ as the chordal flag-median objective function value. Then*

$$f(T([\![\mathbf{Y}]\!])) \leq f([\![\mathbf{Y}]\!]). \tag{9}$$

Proof sketch. We define the function

$$h([\![\mathbf{Z}]\!], [\![\mathbf{Y}]\!]) = \sum_{i=1}^p w_i([\![\mathbf{Z}]\!]) d_c([\![\mathbf{X}^{(i)}]\!], [\![\mathbf{Y}]\!])^2. \tag{10}$$

By definition of h, T, and f, we have

$$h(T([\![\mathbf{Y}]\!]), [\![\mathbf{Y}]\!]) \leq h([\![\mathbf{Y}]\!], [\![\mathbf{Y}]\!]) \leq f([\![\mathbf{Y}]\!]).$$

We use h and $2a - b < \frac{a^2}{b}$ for $a, b \in \mathbb{R}, b > 0$ to find

$$2f(T([\![\mathbf{Y}]\!]) - f([\![\mathbf{Y}]\!]) \leq h(T([\![\mathbf{Y}]\!]), [\![\mathbf{Y}]\!]).$$

From our string of inequalities, we have the desired result. We leave the full proof to our supplementary material. □

Remark 3. *The distance vanishes when $[\![\mathbf{Y}]\!] = [\![\mathbf{X}^{(i)}]\!]$ (e.g., $d_c([\![\mathbf{Y}]\!], [\![\mathbf{X}^{(i)}]\!]) = 0$). In this case, Alg. 2 gets stuck at $[\![\mathbf{X}^{(i)}]\!]$ and the result in Prop. 5 becomes*

$$f(T([\![\mathbf{Y}]\!])) \leq f([\![\mathbf{Y}]\!]) + p\epsilon/2. \tag{11}$$

This singularity can be removed even for a general Weiszfeld iteration, simply by replacing the weights [5].

Proposition 6. *Let $[\![\mathbf{Y}_k]\!] \in \mathcal{FL}(d+1)$ be an iterate of Alg. 2 and $f : \mathcal{FL}(d+1) \to \mathbb{R}$ denote the chordal flag-median objective value. $f([\![\mathbf{Y}_k]\!])$ converges as $k \to \infty$ as long as $d_c([\![\mathbf{Y}]\!], [\![\mathbf{X}_i]\!]) > \epsilon$ for $i = 1, 2, \dots, p$ and each k.*

Proof. Notice that the real sequence with terms $f([\![\mathbf{Y}_k]\!]) \in \mathbb{R}$ is bounded below by 0 and is decreasing by Prop. 5. So it converges as $k \to \infty$. □

4. Motion Averaging

In this section, we propose a method for motion averaging by leveraging novel definitions of averages on the flag manifold. This will also act as a good example of how to use flag manifolds for performing computations on other groups. To this end, we now define the group of 3D rotations and translations, $SE(3)$. Then we outline how to navigate between points on $SE(3)$ and points on a flag. Finally, we describe our *motion averaging on flag manifolds*.

Definition 9 (3D motion)**.** *The configuration (position and orientation) of a rigid body moving in free space can be described by a homogeneous transformation matrix* $\mathbf{M}$ *corresponding to the displacement from any inertial reference frame to another. The set of all such rigid body transformations in three-dimensions form the $SE(3)$ group:*

$$SE(3) = \left\{ \boldsymbol{\gamma} := \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} : \mathbf{R} \in SO(3) \text{ and } \mathbf{t} \in \mathbb{R}^3 \right\},$$

where $\mathbf{t}$ *denotes a* translation *(positional displacement) and* $\mathbf{R}$ *captures the angular displacements as an element of the special orthogonal group $SO(3)$:*

$$SO(3) = \left\{ \mathbf{R} \in \mathbb{R}^{3\times 3} : \mathbf{R}^\top \mathbf{R} = \mathbf{I} \wedge \det \mathbf{R} = 1 \right\}. \tag{12}$$

Proposition 7 (Motion contraction [43])**.** *We call* $\Phi_\lambda : SE(3) \to SO(4)$ *a* Saletan contraction, i.e. $\Phi_\lambda(\boldsymbol{\gamma}) = \mathbf{U}\mathbf{V}^T$ *where the left ($\mathbf{U}$) & right ($\mathbf{V}$) singular vectors are obtained via the singular value decomposition:*

$$\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \begin{bmatrix} \mathbf{R} & \mathbf{t}/\lambda \\ \mathbf{0}^\top & 1 \end{bmatrix} \text{ for } \boldsymbol{\gamma} \in SE(3). \tag{13}$$

Proposition 8 (Inverse motion contraction [43])**.** *We call the inverse contraction map $\Phi_\lambda^{-1} : SO(4) \to SE(3)$. Let $\mathbf{M} \in SO(4)$, then $\boldsymbol{\gamma} = \Phi_\lambda^{-1}(\mathbf{M})$ where*

$$\mathbf{t} = \frac{2\lambda}{\mathbf{M}_{4,4}} \mathbf{M}_{1:3,4}, \tag{14}$$

$$\mathbf{R} = \begin{cases} \mathbf{M}_{1:k,1:k}, & \|\mathbf{t}\|_2 < \epsilon \\ \left(\mathbf{M}_{4,4} \frac{\mathbf{t}\mathbf{t}^T}{\|\mathbf{t}\|_2^2} + \mathbf{P}'\right)^{-1} \mathbf{M}_{1:k,1:k}, & \text{o.w.} \end{cases}, \tag{15}$$

and $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathbf{T} = \mathbf{t}^\mathbf{T}$ is the SVD and $\mathbf{P}' = \mathbf{V}_{:,2:4}\mathbf{V}^T_{:,2:4}$.

Algorithm 3: Motion averaging on Flag manifolds.

Input: Motions $\{\gamma\}_{i=1}^p \subset SE(3)$, scale $\lambda \in R$
Output: Average motion $\boldsymbol{\gamma}^* \in SE(3)$
Compute $\{\Phi_\lambda(\gamma_i)\}_{i=1}^p \subset SO(4)$ using Prop. 7
Compute $\left\{[\![\mathbf{X}^{(i)}]\!]\right\}_{i=1}^p \subset FL^+(1,2,3;4)\}$ from $\{\Phi(\gamma_i)\}_{i=1}^p$ using Prop. 9
Mean: $[\![\mathbf{Y}^*]\!] \leftarrow$ flag-mean $\left(\left\{[\![\mathbf{X}^{(i)}]\!]\right\}_{i=1}^p\right)$
Median: $[\![\mathbf{Y}^*]\!] \leftarrow$ flag-median $\left(\left\{[\![\mathbf{X}^{(i)}]\!]\right\}_{i=1}^p\right)$
Use Prop. 10 to compute $\mathbf{M}^\star \in SO(4)$
Use Prop. 8 to compute $\mathbf{R}^\star \in SO(3)$ and $\mathbf{t}^\star \in \mathbb{R}^3$

Proposition 9 (Flag representation of motion [45])**.** *Any contracted motion $\mathbf{M} \in SO(4)$ can be represented as a point on the flag, $[\![\mathbf{X}]\!] \in \mathcal{FL}^+(1,2,3;4)$ as the first 3 columns of $\mathbf{M}$. Namely, $[\![\mathbf{X}]\!]$ is*

$$[\mathbf{m_1}] \subset [\mathbf{m_1}, \mathbf{m_2}] \subset [\mathbf{m_1}, \mathbf{m_2}, \mathbf{m_3}] \subset \mathbb{R}^4. \tag{16}$$

Remark 4. *Note that the elements of the group of rigid body motions, $SE(3)$, which we represent by points on $SO(4)$, can be imagined as the points of a six-dimensional quadric in seven-dimensional projective space, $\mathbb{P}^7$, called the* Study quadric *[45]. The well known dual quaternions are the very coordinates of this space. Such a bijection between $\mathbb{P}^7$ and $SO(4)$ [36] is the reason why our free parameter λ resembles the* dual unit *ε in dual quaternions [45, 1, 13]. Moreover, our flag manifold, $\mathcal{FL}^+(1,2,3;4)$ is homeomorphic to $SO(4)$. We leave the investigation of these deeper connections to future work.*

Proposition 10 (Motion representation of a flag [45])**.** *Given $[\![\mathbf{X}]\!] \in \mathcal{FL}^+(1,2,3;4)$ with the same basis vectors from Prop. 9, the corresponding point on $SO(4)$ is*

$$[\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \mathbf{z}] \in SO(4), \tag{17}$$

where $\mathbf{z}$ is found by running the Gram-Schmidt process to find a 4^{th} unit vector orthogonal to span$\{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}$.

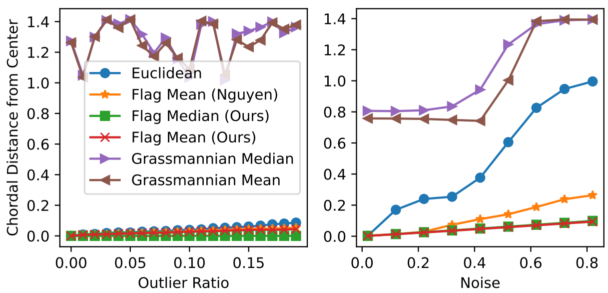

Figure 2: 100 points from a synthetic data set on $\mathcal{FL}(1,3;10)$. The vertical axis is the chordal distance on $\mathcal{FL}(1,3;10)$ between the predicted averages and the "center" of the data set.

4.1. Single Motion Averaging

With these constructs, we are now ready to formally define the motion averaging problem for points on $SE(3)$.

Definition 10. *Given a set of motions $\{\boldsymbol{\gamma}_i \in SE(3)\}_{i=1}^p$, the centroid is defined to be the solution of the following optimization procedure:*

$$\boldsymbol{\gamma}^* = \underset{\boldsymbol{\gamma} \in SE(3)}{\arg\min} \sum_{i=1}^{p} \alpha_i \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}\|_{\mathrm{F}}^q \tag{18}$$

where $q = 2$ for mean estimation, $q = 1$ for the median and $\alpha_i \in \mathbb{R}$ denote the individual weights.

To solve Eq. 18, we simply map each $\boldsymbol{\gamma}_i \in SE(3)$ to $\mathbf{X}^{(i)} \in FL(1,2,3;4)^+$. To this end, we first map each $\boldsymbol{\gamma}_i$ to $\phi_\lambda(\boldsymbol{\gamma}_i) = \mathbf{M}_i \in SO(4)$ via Prop. 7 and subsequently use Prop. 9 to represent $\mathbf{M}_i$ as $[\![\mathbf{X}^{(i)}]\!] \in FL(1,2,3;4)^+$. Then we use our flag-mean ($q = 2$) or -median algorithm ($q = 1$) to solve

$$[\![\mathbf{Y}^*]\!] = \underset{[\![\mathbf{Y}]\!] \in FL(1,2,3;4)^+}{\arg\min} \sum_{i=1}^{p} \alpha_i d_c([\![\mathbf{X}^{(i)}]\!], [\![\mathbf{Y}]\!])^q \tag{19}$$

The desired solution $\boldsymbol{\gamma}^* \in SE(3)$ is then obtained by first mapping $[\![\mathbf{Y}^*]\!]$ back to $\mathbf{M}^* \in SO(4)$ via Prop. 10 and subsequently using $\boldsymbol{\gamma}^* = \phi_\lambda^{-1}(\mathbf{M}^*)$ by Prop. 8. We present this chordal Flag motion averaging in Alg. 3.

5. Results

5.1. Averaging on Flag Manifolds

We first consider examples of data naturally existing as flags. We work with 5 data sets: 2 synthetic ones, MNIST digits [16], the Yale Face Database [9], and the Cats and Dogs dataset [48]. We provide further evaluation of our flag averages that result in improved clustering on the UFC YouTube dataset [28] in the supplementary material. In one synthetic experiment, we compare our Stiefel Riemannian Trust-Regions (RTR) method in Alg. 1 for computing the flag-mean to the Flag RTR by Nguyen *et al.* [37]. In the rest of the experiments, we compare our *chordal* flag (FL)-mean & -median to the Grassmannian (GR)-mean [18] & -median [31], as well as Euclidean averaging, where the matrices are simply averaged and projected onto the flag manifold via QR decomposition. GR-means and -medians, [18, 31] input data a points on Grassmannians by using the largest dimensional subspace in the flag ($[\mathbf{X}^{(i)}] \in \mathrm{Gr}(k,d)$) and output an average as a flag of type $(1,2,\dots,k,d)$. So all the methods considered in this section result in averages which live on a flag manifold. In this section we compare methods for data representation: the flag vs. Grassmannian vs. Euclidean space.

	Dist. to $\mathbf{C}$	Obj. Fn. Value
Ours	$(1.4 \pm 0.2) \times 10^{-4}$	$(2.1 \pm 0.05) \times 10^{-4}$
[37]	$(3.0 \pm 2.1) \times 10^{-3}$	$(1.6 \pm 1.6) \times 10^{-3}$

Table 1: Robustness to initialization: Alg. 1 versus Flag RTR from Nguyen *et al.* [37]. Data: 100 points on $\mathcal{FL}(1,2,3;10)$.

Synthetic data. Both our synthetic experiments use the same methodology for generating data sets on the Grassmannian and flag. We begin by computing a "center" representative, $\mathbf{C} \in \mathbb{R}^{10\times 3}$, as the first 3 columns of the QR decomposition of a random matrix in $\mathbb{R}^{10\times 3}$ with entries sampled from the uniform distribution over $[-.5, .5)$, $\mathcal{U}[-.5,.5)$. The representative for the i^{th} data point, $\mathbf{X}_i$, is computed by sampling $\mathbf{Z}_i \in \mathbb{R}^{10\times 3}$ with entries from $\mathcal{U}[-.5,.5)$ and defined as the first 3 columns of the QR decomposition of $\mathbf{C} + \delta\mathbf{Z}_i$ for a noise parameter $\delta \geq 0$.

Averaging synthetic flag data. We use synthetic data sets with 100 points, on $\mathrm{Gr}(3;10)$ and $\mathcal{FL}(1,3;10)$. For the left plot in Fig. 2 we vary δ to compute our data sets. For the right plot we have m outliers computed with $\delta = 1$ and the rest of the data are computed with $\delta = 0.001$. We compute the error as the chordal distance on $\mathcal{FL}(1,3;10)$ between the predicted average and $[\![\mathbf{C}]\!]$. In addition to comparing our averages to Grassmannian (GR) averages, we compare Alg. 1 to Nguyen *et al.* [37] for computing the flag-mean. Our results indicate that our algorithm improves both upon GR, Euclidean, and Nguyen *et al.* [37] averages in the sense that flag averages are closer to $[\![\mathbf{C}]\!]$. Specifically, our flag-median is more robust to outliers than our flag-mean. Note: Euclidean out preforms GR averaging because Euclidean averaging respects column order (e.g., the flag structure) for matrix representatives of the data, whereas GR averaging does not.

Comparisons to Riemannian flag optimization. In a second experiment, we compare the convergence of Alg. 1 to that of Flag RTR [37]. To this end, we generate 100 points

Figure 3: Averaging a collection of faces belonging to three different people, captured under varying illumination: center, left and right. Notice that the first dimension of the flag representations is center illuminated, better representing the mean compared to Grassmannian.

on $\mathcal{FL}(1,2,3;10)$ using $\delta = 0.001$ and run 50 random trials with different initializations and compute 3 items (i) the number of iterations to convergence, (ii) the chordal distance on $\mathcal{FL}(1,2,3;10)$ between the flag-mean and $[\![\mathbf{C}]\!]$, (iii) the cost function values from Eq. 2. We find that in every experiment Alg. 1 converges in *2 iterations* and Flag RTR converges, on average, in 9.74 ± 2.76 iterations. In Tab. 1 we see that our method is one order of magnitude closer to the ground truth centroid $[\![\mathbf{C}]\!]$ and produces a one order of magnitude smaller objective function value.

Averaging under varying illumination. To further demonstrate the efficacy of our averages over the standard Grassmanians, we leverage face images from Yale Face Database [9] with central (c), left (l), and right (r) illuminations, respectively. Let $\mathbf{A_c}, \mathbf{A_l}, \mathbf{A_r} \in \mathbb{R}^{243\times 320}$ be these three images of a person. We represent a face as a point $[\![\mathbf{X}]\!] \in \mathcal{FL}(1,3;d)$ as $[\![\mathbf{X}]\!] = [\mathbf{X}_1] \subset [\mathbf{X}] \subset \mathbb{R}^d$ and as $[\mathbf{X}] \in \mathrm{Gr}(3,d)$ using the following three steps: (i) Set $\mathbf{v_i} = \mathrm{vec}(\mathbf{A_i})$ for $i = c,l,r$; (ii) take $\mathbf{X} = \mathbf{Q}_{:,1:3}$ where $\mathbf{Q}$ is from the QR decomposition of $[\mathbf{v_c}, \mathbf{v_l}, \mathbf{v_r}]$. Repeating this process for three faces gives us three points: $[\mathbf{X}_1], [\mathbf{X}_2], [\mathbf{X}_3] \in \mathrm{Gr}(3,d)$ and $[\![\mathbf{X}_1]\!], [\![\mathbf{X}_2]\!], [\![\mathbf{X}_3]\!] \in \mathcal{FL}(1,3;d)$. Then we calculate the Grassmannian-mean of the points in $\mathrm{Gr}(3,d)$ which is the flag: $[\![\boldsymbol{\nu}]\!] = [\boldsymbol{\nu}_1] \subset [\boldsymbol{\nu}_1, \boldsymbol{\nu}_2] \subset [\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3]$ and the flag-mean (ours) of the points in $\mathcal{FL}(1,3;d)$: $[\![\boldsymbol{\mu}]\!] = [\boldsymbol{\mu}_1] \subset [\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3]$. A plot of reshaped $\boldsymbol{\mu}_1$ and $\boldsymbol{\nu}_1$ for a set of three faces in Fig. 3. We would expect the first dimension of both means to look like a face with center illumination. However, only the flag-mean appears to be center-illuminated.

MNIST representation. We run two experiments similar to what was done in [31] with MNIST digits. However, our representations differ since we represent a digit as $[\mathbf{X}_j] \in \mathrm{Gr}(2,784)$ and $[\![\mathbf{X}_j]\!] \in \mathcal{FL}(1,2;784)$. We generate p representations of a digit, $\{\mathbf{X}_j\}_{j=1}^p$, by sampling a set of p images without replacement from the test partition. Then we vectorize each image into $\mathbf{v}_j \in \mathbb{R}^{784}$ and run k nearest neighbors on $\{\mathbf{v_j}\}_{i=1}^p$ with $k = 2$ using the cosine distance. Say $\mathbf{v}_j$ and $\mathbf{v}_k$ are the 2 nearest neighbors of $\mathbf{v}_j$, then the representation for sample j is $\mathbf{X}_j = \mathbf{Q}_{:,:2}$ from the QR decomposition of $[\mathbf{v}_j, \mathbf{v}_k]$.

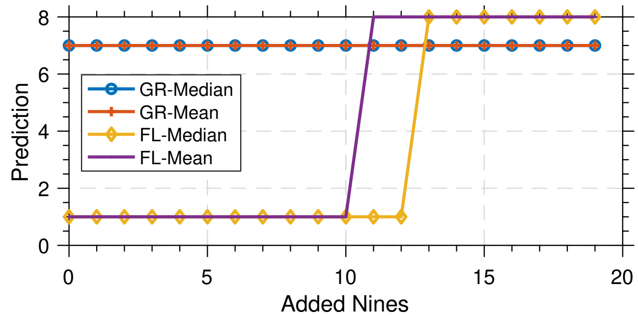

Figure 4: Neural network predictions for the first dimension of different averages $i = 0,1,2,\ldots,19$ MNIST data sets. The ith data set has i representations of the 9s digit and 20 representations of the 1s digit.

Robustness to Neural Network (NN) predictions. For the first MNIST experiment, we use the method above to create 20 data sets on $\mathrm{Gr}(2,784)$ and $\mathcal{FL}(1,2;784)$ corresponding to $i = 0,1,2,\ldots,19$. The ith data set contains 20 representations of the digit 1 and i representations for the digit 9. We calculate a GR-mean and -median of each of the i data sets on $\mathrm{Gr}(2,784)$ and our flag-mean and -median for the data sets on $\mathcal{FL}(1,2,784)$. Note: all of these averages live on $\mathcal{FL}(1,2,784)$. We then use a NN (trained on the original training data and producing a 97% test accuracy on the original test data) to predict the label of the first dimension of each average for $i = 0,1,2,\ldots,19$. As plotted in Fig. 4, the NN incorrectly predicts the class of the GR-mean and -median for each data set. In contrast, the flag-mean and -median are all predicted as 1s with data sets with fewer than 11 representations of the 9s digits. The flag-mean is the first flag average to be incorrectly predicted, since it is not as robust to outliers as the flag-median.

Visualizing robustness. Our second MNIST experiment is with 20 representations of 6s and with i outlier representations of 7s for $i = 0,4,8,12$. We use the workflow from Fig. 4 to represent the MNIST digits on $\mathrm{Gr}(2,748)$ and $\mathcal{FL}(1,2;748)$. For each i, we compute averages of a data set with i representations of 7s. A chordal distance matrix on $\mathcal{FL}(1,2;798)$ between all the averages and data is used to preform Multidimensional Scaling (MDS) [26] for visualization in Fig. 5. The best averages should barely move (right to left) as we add outlier representations of 7s. Our flag-mean and -median are moved the least with the addition of representations of 7s with the median moving less than the mean. In contrast, the Grassmannian-mean and -median [31] move more than the compared baselines as we add 7s.

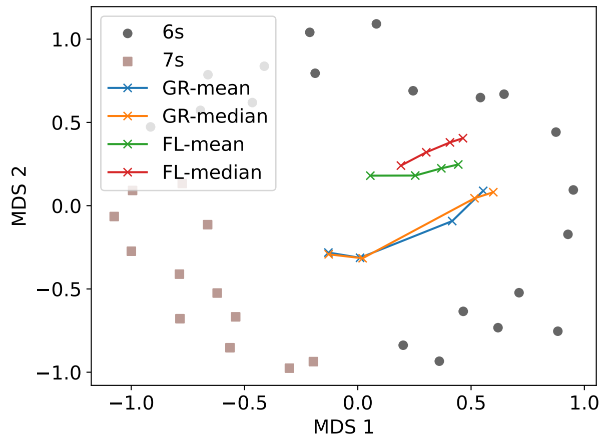

Figure 5: MDS embedding of MNIST digits and Grassmannian and flag averages. Each "x" is an average of 20 representations of 6s as we gradually add i outlier representations of 7s for $i = 0, 4, 8, 12$ data sets. The averages move from right to left as we add more 7s.

PCA by flag statistics. We use the Cats and Dogs dataset [48] to compute 3-dimensional PCA [24] weights, $\mathbf{W}^* \in \mathbb{R}^{4096\times 3}$, of the data matrix, $\mathbf{X} \in \mathbb{R}^{198\times 4096}$. Then we randomly split the m subjects into p evenly sized groups to generate p data matrices each of size p_i: $\{\mathbf{X_i}\}_{i=1}^{p} \subset \mathbb{R}^{p_i\times 4096}$. PCA weights of each $\mathbf{X_i}$ are computed as $\mathbf{W_i} \in \mathbb{R}^{4096\times 3}$. $\mathbf{W}^*$ is predicted by averaging $\{\mathbf{W_i}\}_{i=1}^{p}$ as points on $\mathcal{FL}(1,2,3;4096)$ and $\mathrm{Gr}(3;4096)$. Specifically, we compute the flag-mean (ours), Grassmannian-mean, Euclidean-mean, and a random point. Then we record the chordal distance on $\mathcal{FL}(1,2,3;4096)$ (reconstruction error) between the average and $[\![\mathbf{W}^*]\!] \in \mathcal{FL}(1,2,3;4096)$. Our flag-mean is closer to $[\![\mathbf{W}^*]\!]$ for $p = 1,2,\dots,6$.

5.2. Averaging Rigid Motions

We now evaluate our algorithm in robust averaging of a set of points represented on the $SE(3)$-manifold. To this end, we synthesize a dataset of 400 rigid motions (rotations and translations) around multiple randomly drawn central points in $SE(3)$. These points are generated with increasing noise levels. Particularly, for rotations we perturb the rotation axis using variances of $[0, 5, 10, 15, 20, 25]$ degrees, while the translations are perturbed in the levels of $[0, 0.02, 0.05, 0.1, 0.2, 0.3]$. For each noise level, we run 50 experiments and use $\lambda = 1$ to ensure that translations and rotations are well balanced. We then run our algorithms for the flag-mean and -median. These algorithms are compared to standard Govindu [20], and baseline (QT) where translations and quaternions are averaged independently using Markley's method [32]. We also ran dual quaternion averaging of Torsello *et al*. [46] and found it produced identical results to Govindu. Our results in Fig. 7 show that both of our algorithms surpass classical motion averages with our flag-median producing more robust estimates.

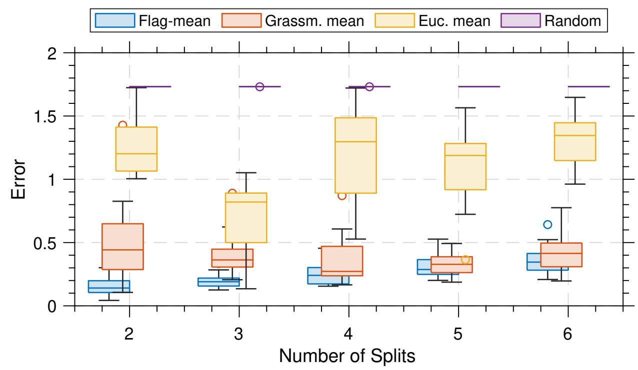

Figure 6: Reconstruction error for PCA weights as a function of Number of Splits, p.

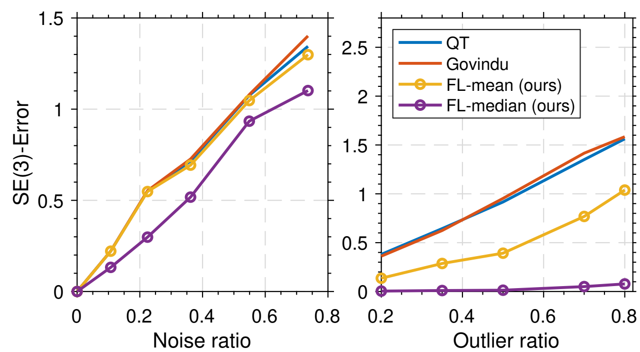

Figure 7: Single motion averaging experiments for increasing levels of $SE(3)$-noise and outlier ratios.

6. Conclusion

We have provided two algorithms, the *flag-mean* & *flag-median*, that estimate flag-prototypes of points defined on flag manifolds using chordal distance. We have established the convergence of our IRLS algorithm yielding the flag-median. Our methodologies deviate from the existing literature [18, 31] which average Grassmannians into flags, and are found to be useful when either inherent outlier-robustness is necessary or when the subspaces possess a natural order, (e.g., hierarchical data). Since flag manifolds generalize Grassmannians, our methods can average on a broader class of manifolds. Consequently, we have applied our averages to rigid motions via group contraction.

Limitations & future work. Our method can become computationally expensive when applied to high-dimensional problems. Moreover, our convergence results are weaker than desired as we have not provided a convergence rate. Besides addressing these, our future work involves clustering and inference on data with hierarchical structures.

Acknowledgements. Benjamin Busam introduced Nathan and Tolga during CVPR 2022 in New Orleans. This gracious act is the catalyst in the realization of this work.

References

[1] Rafal Ablamowicz, Garret Sobczyk, et al. *Lectures on Clifford (geometric) algebras and applications*. Springer, 2004. 5

[2] P-A Absil, Christopher G Baker, and Kyle A Gallivan. Trust-region methods on Riemannian manifolds. *Foundations of Computational Mathematics*, 7(3):303–330, 2007. 3

[3] Bijan Afsari. Riemannian $\mathcal{L}^p$ center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011. 3

[4] Khurrum Aftab and Richard Hartley. Convergence of iteratively re-weighted least squares to robust m-estimators. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 480–487. IEEE, 2015. 4

[5] Khurrum Aftab, Richard Hartley, and Jochen Trumpf. Generalized weiszfeld algorithms for lq optimization. *IEEE transactions on pattern analysis and machine intelligence*, 37(4):728–745, 2014. 5

[6] Giovanni Alberti. Geometric measure theory. *Encyclopedia of mathematical Physics*, 2:520–527, 2005. 3

[7] DV Alekseevsky. Flag manifolds. *Sbornik Radova*, 11:3–35, 1997. 2

[8] Amir Beck and Shoham Sabach. Weiszfeld's method: Old and new results. *Journal of Optimization Theory and Applications*, 164:1–40, 2015. 4

[9] Peter N. Belhumeur, Joao P Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):711–720, 1997. 6, 7

[10] J Ross Beveridge, Bruce A Draper, Jen-Mei Chang, Michael Kirby, Holger Kley, and Chris Peterson. Principal angles separate subject illumination spaces in ydb and cmu-pie. *IEEE transactions on pattern analysis and machine intelligence*, 31(2):351–363, 2008. 1

[11] Tolga Birdal and Umut Simsekli. Probabilistic permutation synchronization using the riemannian structure of the birkhoff polytope. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11105–11116, 2019. 1

[12] Nicolas Boumal, Bamdev Mishra, P-A Absil, and Rodolphe Sepulchre. Manopt, a Matlab toolbox for optimization on manifolds. *The Journal of Machine Learning Research*, 15(1):1455–1459, 2014. 1, 3

[13] Benjamin Busam, Tolga Birdal, and Nassir Navab. Camera pose filtering with local regression geodesics on the Riemannian manifold of dual quaternions. In *IEEE International Conference on Computer Vision Workshop (ICCVW)*, October 2017. 5

[14] Rudrasis Chakraborty, Soren Hauberg, and Baba C Vemuri. Intrinsic grassmann averages for online linear and robust subspace learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6196–6204, 2017. 1

[15] Rudrasis Chakraborty and Baba C Vemuri. Recursive frechet mean computation on the grassmannian and its applications to computer vision. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4229–4237, 2015. 1

[16] Li Deng. The MNIST database of handwritten digit images for machine learning research. *IEEE signal processing magazine*, 29(6):141–142, 2012. 6

[17] Ron Donagi and Eric Sharpe. Glsms for partial flag manifolds. *Journal of Geometry and Physics*, 58(12):1662–1692, 2008. 2

[18] Bruce Draper, Michael Kirby, Justin Marks, Tim Marrinan, and Chris Peterson. A flag representation for finite collections of subspaces of mixed dimensions. *Linear Algebra and its Applications*, 451:15–32, 2014. 3, 6, 8

[19] P Thomas Fletcher and Sarang Joshi. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, 87(2):250–262, 2007. 1

[20] Venu Madhav Govindu. Lie-algebraic averaging for globally consistent motion estimation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2004.*, volume 1, pages I–I. IEEE, 2004. 8

[21] Mehrtash T Harandi, Conrad Sanderson, Sareh Shirazi, and Brian C Lovell. Graph embedding discriminant analysis on grassmannian manifolds for improved image set matching. In *CVPR 2011*, pages 2705–2712. IEEE, 2011. 1

[22] Jun He, Laura Balzano, and Arthur Szlam. Incremental gradient on the grassmannian for online foreground and background separation in subsampled video. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1568–1575. IEEE, 2012. 1

[23] Yi Hong, Roland Kwitt, Nikhil Singh, Brad Davis, Nuno Vasconcelos, and Marc Niethammer. Geodesic regression on the grassmannian. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part II 13*, pages 632–646. Springer, 2014. 1

[24] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933. 8

[25] Michael Kirby. *Geometric data analysis: an empirical approach to dimensionality reduction and the study of patterns*, volume 31. Wiley New York, 2001. 2

[26] Joseph B Kruskal and Myron Wish. *Multidimensional scaling*, volume 11. Sage, 1978. 7

[27] Sriram Kumar and Andreas Savakis. Robust domain adaptation on the l1-grassmannian manifold. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 103–110, 2016. 1

[28] Jingen Liu, Jiebo Luo, and Mubarak Shah. Recognizing realistic actions from videos "in the wild". In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1996–2003. IEEE, 2009. 6

[29] Xiaofeng Ma, Michael Kirby, and Chris Peterson. The flag manifold as a tool for analyzing and comparing sets of data sets. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4185–4194, 2021. 1, 2

[30] Xiaofeng Ma, Michael Kirby, and Chris Peterson. Self-organizing mappings on the flag manifold with applications to hyper-spectral image data analysis. *Neural Computing and Applications*, 34(1):39–49, 2022. 2

[31] Nathan Mankovich, Emily J King, Chris Peterson, and Michael Kirby. The flag median and FlagIRLS. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10339–10347, 2022. 2, 3, 4, 6, 7, 8

[32] F Landis Markley, Yang Cheng, John L Crassidis, and Yaakov Oshman. Averaging quaternions. *Journal of Guidance, Control, and Dynamics*, 30(4):1193–1197, 2007. 8

[33] Timothy Marrinan, J Ross Beveridge, Bruce Draper, Michael Kirby, and Chris Peterson. Flag-based detection of weak gas signatures in long-wave infrared hyperspectral image sequences. In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXII*, volume 9840, pages 407–416. SPIE, 2016. 2

[34] Tim Marrinan, J Ross Beveridge, Bruce Draper, Michael Kirby, and Chris Peterson. Finding the subspace mean or median to fit your need. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1082–1089, 2014. 1, 2

[35] Breton Lawrence Minnehan. *Deep Grassmann Manifold Optimization for Computer Vision*. Rochester Institute of Technology, 2019. 1

[36] Georg Nawratil. Fundamentals of quaternionic kinematics in Euclidean 4-space. *Advances in Applied Clifford Algebras*, 26(2):693–717, 2016. 5

[37] Du Nguyen. Closed-form geodesics and optimization for Riemannian logarithms of Stiefel and flag manifolds. *Journal of Optimization Theory and Applications*, pages 1–25, 2022. 2, 6

[38] Yasunori Nishimori, Shotaro Akaho, Samer Abdallah, and Mark D Plumbley. Flag manifolds for subspace ICA problems. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, volume 4, pages IV–1417. IEEE, 2007. 2

[39] Yasunori Nishimori, Shotaro Akaho, and Mark D Plumbley. Riemannian optimization method on generalized flag manifolds for complex and subspace ICA. In *AIP Conference Proceedings*, volume 872, pages 89–96. American Institute of Physics, 2006. 2

[40] Yasunori Nishimori, Shotaro Akaho, and Mark D Plumbley. Riemannian optimization method on generalized flag manifolds for complex and subspace ica. In *AIP Conference Proceedings*, volume 872, pages 89–96. American Institute of Physics, 2006. 2

[41] Yasunori Nishimori, Shotaro Akaho, and Mark D Plumbley. Riemannian optimization method on the flag manifold for independent subspace analysis. In *International conference on independent component analysis and signal separation*, pages 295–302. Springer, 2006. 2

[42] Yasunori Nishimori, Shotaro Akaho, and Mark D Plumbley. Natural conjugate gradient on complex flag manifolds for complex independent subspace analysis. In *International Conference on Artificial Neural Networks*, pages 165–174. Springer, 2008. 2

[43] Onur Ozyesil, Nir Sharon, and Amit Singer. Synchronization over Cartan motion groups via contraction. *SIAM Journal on Applied Algebra and Geometry*, 2(2):207–241, 2018. 1, 5

[44] Renaud-Alexandre Pitaval and Olav Tirkkonen. Flag orbit codes and their expansion to Stiefel codes. In *2013 IEEE Information Theory Workshop (ITW)*, pages 1–5. IEEE, 2013. 2

[45] JM Selig. The study quadric. *Geometric Fundamentals of Robotics*, pages 241–269, 2005. 2, 5

[46] Andrea Torsello, Emanuele Rodola, and Andrea Albarelli. Multiview registration via graph diffusion of dual quaternions. In *CVPR 2011*, pages 2441–2448. IEEE, 2011. 8

[47] Mark Wiggerman. The fundamental group of a real flag manifold. *Indagationes Mathematicae*, 9(1):141–153, 1998. 2

[48] Wendy S Yambor. Analysis of PCA-based and Fisher discriminant-based image recognition algorithms. Master's thesis, Citeseer, 2000. 6, 8

[49] Ke Ye, Ken Sze-Wai Wong, and Lek-Heng Lim. Optimization on flag manifolds. *Mathematical Programming*, 194(1):621–660, 2022. 2, 4

[50] Jiayao Zhang, Guangxu Zhu, Robert W Heath Jr, and Kaibin Huang. Grassmannian learning: Embedding geometry awareness in shallow and deep learning. *arXiv preprint arXiv:1808.02229*, 2018. 1

[51] Yongheng Zhao, Tolga Birdal, Jan Eric Lenssen, Emanuele Menegatti, Leonidas Guibas, and Federico Tombari. Quaternion equivariant capsule networks for 3d point clouds. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pages 1–19. Springer, 2020. 4